J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 9, No.1, 23-32, Juli 2012

ISBN:1858-1382

Perbandingan Metode Fraksi dengan Himpunan Bagian Terbaik dalam Pemilihan Model Regresi Berganda

Arlita Aristianingsih Jufra[1], Raupong[1], Anisa[1]

Abstrak

Model regresi terbaik adalah model yang dapat menjelaskan perilaku peubah tak bebas dengan baik dengan memilih peubah-peubah bebas dari sekian banyak peubah bebas yang tersedia dalam data. Dalam pemilihan model regresi terbaik dipilih model yang mempunyai kesalahan prediksi paling sedikit dan melibatkan peubah bebas sesedikit mungkin. Metode yang umum digunakan dalam pemilihan model regresi terbaik yaitu Regresi Bertahap, Seleksi Maju, Himpunan Bagian Terbaik dan Metode Fraksi. Berbeda dengan Himpunan Bagian Terbaik, Metode Fraksi digunakan dalam penentuan model regresi terbaik dengan data yang memiliki tingkat multikolinearitas yang tinggi. Contoh penerapan metode Fraksi dan metode Himpunan Bagian Terbaik dilakukan pada data Indeks Pembangunan Manusia, dimana diperoleh bahwa model regresi dengan menggunakan metode Fraksi lebih baik dibandingkan dengan model regresi dengan menggunakan metode HImpunan Bagian Terbaik, dimana pada model regresi dengan menggunakan metode Himpunan Bagian Terbaik masih memiliki multikolinearitas.

Kata Kunci: Model regresi terbaik, himpunan bagian terbaik, metode fraksi, multikolinearitas.

1. Pendahuluan

Pada regresi berganda sering muncul masalah multikolinear atau adanya hubungan linear antar peubah bebas (independen). Masalah multikolinear ditunjukkan oleh nilai *Variance Inflation Factors (VIF)* yang tinggi. Jika terdapat masalah multikolinear maka kesimpulan yang diperoleh dari hasil pengujian untuk model regresi maupun untuk masing-masing peubah yang ada dalam model, seringkali tidak tepat, sehingga diperlukan suatu metode yang dapat memberikan model terbaik dengan peubah-peubah independen yang paling sesuai untuk menjelaskan peubah dependennya. Model regresi terbaik adalah model yang dapat menjelaskan perilaku peubah tak bebas dengan baik dengan memilih peubah-peubah bebas dari sekian banyak peubah bebas yang tersedia dalam data. Misalkan akan ditentukan suatu persamaan regresi linear peubah respon tertentu Y terhadap peubah-peubah bebas atau peramal $X_1, X_2, \ldots, X_k$. Untuk menentukan peubah bebas mana yang akan dimasukkan kedalam model regresi, ada dua kriteria yang saling bertentangan :

1. Agar persamaannya bermanfaat bagi tujuan peramalan, biasanya memasukkan sebanyak mungkin X sehingga diperoleh nilai ramalan yang terandalkan.

[1] *Jurusan Matematika FMIPA Universitas Hasanuddin, email: raupong@yahoo.com, nkalondeng@gmail.com*

2. Karena untuk memperoleh informasi dari banyak peubah serta pemonitorannya seringkali diperlukan biaya yang tinggi, maka diinginkan persamaan regresinya mencakup sesedikit mungkin X.

Kompromi antara kedua ekstrem itulah yang biasanya disebut pemilihan persamaan regresi terbaik [1].

2. Tinjauan Pustaka

2.1. Analisis Regresi Secara Umum

Analisis regresi menggambarkan sekumpulan teknik statistika yang menjadi dasar pengambilan kesimpulam (inferensia) tentang hubungan antar peubah-peubah yang terukur.Dalam bidang statistika terapan, banyak teknik analisis data modern dikembangkan.Perkembangan berbagai macam teknik analisis data ini didasari pada suatu motivasi untuk menyelesaikan suatu permasalahan khusus.Analisis regresi menggambarkan sekumpulan teknik statistika yang menjadi dasar pengambilan kesimpulam (inferensia) tentang hubungan antar peubah-peubah yang terukur.

Multikolinearitas

Multikolinearitas adalah kondisi terdapatnya hubungan linear atau korelasi yang tinggi antara masing-masing peubah bebas dalam model regresi.

Keeratan hubungan antara antara satu peubah dengan peubah lainnya, biasa disebut dengan koefisisen korelasi yang ditandai dengan r. Koefisien korelasi (r),X_i dan Y_i merupakan taksiran dari korelasi populasi dengan kondisi sampel normal (acak). Koefisien korelasi r dapat diketahui melalui persamaan berikut

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum x_i^2 - (\sum_{i=1}^{n} x_i)^2][\sum y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

Koefisien korelasi dapat bernilai paling kecil -1 dan paling besar 1. Nilai-nilai r lainnya terletak antara -1 dan 1 dengan tanda negatif menyatakan adanya hubungan negatif dan tanda positif menyatakan adanya hubungan positif. Bila nilai r mendekati 1 atau mendekati -1, hubungan antara kedua peubah itu kuat dan dikatakan terdapat korelasi yang tinggi antara keduanya. Nilai $r = 0$ (mendekati nol) menyatakan tidak terdapat hubungan linear antara X_i dan Y_i.

Pemilihan Model Regresi Terbaik

Pemilihan model regresi terbaik adalah penentuan peubah bebas mana yang akan dimasukkan dalam model regresi sehingga model tersebut dapat menjelaskan perilaku peubah tak bebas dengan baik. Dalam pemilihan model regresi terbaik terdapat langkah-langkah yang disarankan sebagai berikut.

1. Menentukan model lengkap yang akan dipertimbangkan.
2. Menentukan kriteria untuk memilih sebuah model.
3. Menentukan strategi untuk menerapkan kriteria.
4. Melakukan analisis khusus.

5. Mengevaluasi model yang telah dipilih.

Setiap langkah menjamin keandalan dan mengurangi pekerjaan yang diperlukan [5].

Menentukan Model Lengkap (*Full Model*)

Model lengkap didefinisikan sebagai model terbesar yang dipertimbangkan dalam setiap langkah dalam proses seleksi model. Semua model yang mungkin lainnya dapat dibuat dengan menghilangkan satu atau lebih peubah bebas dari model lengkap.Sebuah model dapat dibuat dengan menghilangkan peubah-peubah dari model lengkap.

Metode Himpunan Bagian Terbaik

Metode himpunan bagian terbaik lebih baik digunakan daripada strategi seleksi peubah yang lain. Hanya strategi/metode ini yang menjamin untuk menemukan model yang mempunyai R_p^2 terbesar, R_{adj}^2, S^2 terkecil, dan $Cp\ Mallows$. Metode ini tidak selalu dapat digunakan karena jumlah perhitungan menjadi tidak praktis jika banyaknya peubah k dalam model lengkap besar.

2.2. Metode Fraksi

Metode Fraksi merupakan metode pemilihan model terbaik yang menggunakan hasil analisis korelasi sebagai dasar pemilihan peubah yang akan dimasukkan dalam model. Peubah-peubah dikelompokkan berdasarkan kedekatannya secara linear. Dimana peubah-peubah dikelompokkan berdasarkan kategori berkorelasi kuat, nyata dan tidak cukup kuat. Suatu kelompok berisi peubah-peubah yang saling berkorelasi kuat. Model terbaik dibentuk dari peubah-peubah yang mewakili masing-masing kelompok. Tetapi dengan ketentuan bahwa peubah yang dipilih adalah fraksi dari jumlah peubah bebas [2].

Dalam pengelompokkannya untuk berkorelasi nyata dikategorikan pada peubah bebas yang berkorelasi dengan peubah terikat dimana $-1 < \rho_{Y_i X_{ij}} < -0,65$ atau $0,65 < \rho_{Y_i X_{ij}} < 1$, sedangkan untuk berkorelasi kuat dikategorikan pada peubah bebas yang saling berkorelasi dengan salah satu peubah bebas lainnya dengan $-1 < \rho_{X_{ij} X_{ij}} < -0,65$ atau $0,65 < \rho_{X_{ij} X_{ij}} < 1$.

2.3. Ukuran Kesesuaian Model

Koefisien Determinasi Berganda ($\boldsymbol{R_p^2}$)

Koefisien determinasi berganda, R_p^2, adalah ukuran bagian peubah terikat yang dapat dijelaskan secara bersama-sama oleh peubah bebas yang ada di dalam model. Nilai koefisien determinasi berganda R_p^2 didapatkan dari

$$R^2 = \frac{JKR_p}{JKtotal}$$

dengan $JKR_p =$ jumlah kuadrat regresi, $JKtotal$ = jumlah kuadrat total. Dari rumus di atas didapatkan nilai koefisien berganda R_p^2 yaitu

$$R_p^2 = 1 - \frac{JKG_p}{JKtotal}$$

JKG = jumlah kuadrat galat dan p = jumlah peubah bebas dalam model. Penyebut adalah konstan untuk semua regresi yang mungkin. R_p^2 akan terus bertambah seiring bertambahnya peubah bebas yang dimasukkan dalam model. Peubah yang potensial akan ditambahkan dalam model adalah yang memberikan penambahan nilai R_p^2 yang cukup berarti.

Koefisien Determinasi Berganda Terkoreksi ($\boldsymbol{R_{adj}^2}$)

R_{adj}^2 adalah penyesuaian dari R^2. Sehingga koefisien determinasi berganda yang disesuaikan adalah.

$$R_{adj}^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{JKK_p}{JKtotal}$$

dengan n = jumlah amatan. Jika konstan tidak termasuk model bagian, maka $(n-1)$ diganti dengan n. Nilai R_{adj}^2 meningkat jika dan hanya jika nilai $(n-p)JKK_p$ turun, karena $(n-1)JKtotal$ tetap. Model yang baik memiliki R_{adj}^2yang besar. R_{adj}^2tidak pernah dapat menjadi negatif seperti R_p^2 [3].

Kuadrat Tengah Sisa ($\boldsymbol{S^2}$)

Bila jumlah amatannya cukup besar, evaluasi terhadap rata-rata kuadrat tengah sisa untuk setiap kelompok seringkali dapat menunjukkan titik pemisah yang terbaik bagi banyaknya peubah yang sebaiknya disertakan dalam regresi. Nilai kuadrat tengah sisa didapatkan dari

$$S^2 = \frac{JKK_p}{n-p-1}$$

Cp–Mallow

Sebuah statistik lain yang memperoleh ketenaran di tahun-tahun belakangan ini adalah statistik C_p, yang pada awalnya dikemukakan oleh C.L. Mallows. Nilai dugaan yang diperoleh dari persamaan regresi berdasarkan sebagian peubah bebas pada umumnya bias. Untuk menilai kebaikan model digunakan *means square error* (*MSE*) dengan dua perbandingan nilai yaitu variansi dan bias, dari nilai prediksi bukan variansi. Bentuk dari statistik C_p Mallow adalah

$$C_p = \frac{JKS_p}{\hat{\sigma}^2} - (n-2p)$$

yang memperhitungkan baik bias serta variansi dimana, JKS_p adalah jumlah kuadrat sisa dari model yang mengandung parameter p (p adalah jumlah parameter dalam model termasuk β_0) dan $\hat{\sigma}^2 = s^2$ adalah kuadrat tengah sisa dari persamaan terbesar yang dipostulatkan mengandung semua X (yang diasumsikan merupakan nilai dugaan tak bias yang terandalkan bagi ragam galat

σ^2). Cp Mallow berkaitan erat dengan statistik R_p^2 , statistik R^2 terkoreksi (R_{adj}^2), statistik R^2, serta dengan statistik F parsial.

***Variance Inflation Factor (VIF)* dalam Regresi**

Kolinearitas dapat meningkatkan estimasi variansi parameter; model hasil dimana tidak ada peubah yang secara statistik signifikan meskipun R_y^2 besar [1]. Praktisi seringkali tidak tepat menerapkan aturan atau kriteria yang menunjukkan bila nilai *VIF* atau toleransi telah mencapai tingkat yang sangat tinggi. Tidak jarang sebuah *VIF* dengan nilai 10 atau bahkan 1, sama nilainya dengan 4 (setara dengan tingkat toleransi dari 0,10 atau 0,25) telah digunakan sebagai aturan praktis untuk menunjukkan multikolinearitas. Aturan-aturan untuk terlalu banyak multikolinearitas terlalu sering digunakan untuk mempertanyakan hasil analisis yang cukup padat atas dasar statistik [4].

3. Metode Analisis Data

3.1. Jenis dan Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder berupa **Data Indeks Pembangunan Manusia** yang diperoleh dari http:/www.jabarprov.go.id, dimana terdiri dari satu peubah terikat/dependen dan delapan peubah bebas.

Indeks Pembangunan Manusia (IPM) atau *Human Development Index* (HDI) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua Negara di seluruh dunia. Penyusunan basis data IPM bertujuan untuk melihat perkembangan pembangunan manusia, memberi gambaran yang lebih sederhana dan lengkap dalam melihat dampak pembangunan yang dilaksanakan dan implikasinya terhadap peningkatan kualitas penduduk. Juga untuk melihat gambaran tentang seberapa besar kemajuan IPM di masing-masing kabupaten/kota setiap tahunnya dan bagaimana kontribusi kabupaten/kota dalam menunjang akselerasi pencapaian IPM.

3.2. Metode dan Analisis

Data sekunder yang diperoleh dianalisis dengan menggunakan Metode Himpunan Bagian Terbaik dan Metode Fraksi. Dalam kajian ini digunakan data yang mempunyai 8 peubah bebas. Data tersebut mempunyai korelasi tinggi antar peubah bebasnya. Agar terlihat jelas perbandingan dari hasil kedua metode tersebut, penyajian dilakukan untuk data.

Adapun langkah-langkah analisisnya adalah:

1. Menyajikan model maksimum yang akan dipertimbangkan.
2. Memilih model menggunakan Metode Himpunan Bagian Terbaik dan Metode Fraksi.
3. Menentukan kriteria untuk memilih model.
   Sebuah kriteria seleksi adalah sebuah indeks yang dapat dihitung untuk setiap calon model dan digunakan untuk membandingkan beberapa model. Terdapat 5 kriteria yaitu $R_p^2,\ R_{adj}^2, S^2, C_p\ , dan\ VIF$, dimana kajian pada penelitian ini difokuskan pada nilai R_{adj}^2 dan VIF.
4. Melakukan analisis
   Setelah menyajikan model maksimum, menganalisis model dengan menggunakan Metode Himpunan Bagian Terbaik dan Metode Fraksi dan menentukan kriteria untuk

memilih model, kemudian melakukan analisis data dengan menggunakan bantuan komputer, dengan menggunakan software Minitab.

5. Mengevaluasi model yang terpilih.

4. Hasil dan Pembahasan

Pemilihan persamaan regresi terbaik adalah pemilihan persamaan yang mempunyai kesalahan prediksi paling sedikit dan melibatkan peubah bebas sesedikit mungkin. Untuk memilih persamaan penduga terbaik, maka seharusnya setiap persamaan regresi penduga dievaluasi menurut kriteria tertentu yang telah dijelaskan dimuka yaitu :

1. Nilai R^2 maksimum,
2. Nilai R^2_{adj} maksimum, dan
3. Statistik C_pMallows.
4. Nilai *Variance Inflation Factors (VIF).*

4.1. Penaksiran Parameter Regresi Linier Berganda

Metode Kuadrat Terkecil merupakan suatu metode yang paling banyak digunakan untuk menaksir parameter-parameter regresi. Pada model regresi linear berganda juga digunakan metode Kuadrat Terkecil untuk menaksir parameter. Dalam penaksiran parameter dengan menggunakan metode Kuadrat Terkecil diperlukan asumsi-asumsi yang disebut dengan asumsi linear klasik.

Asumsi lain yang digunakan adalah multikonearitas yaitu terjadi korelasi yang kuat antara peubah bebas. Seperti yang telah dijelaskan untuk uji multikolinearitas digunakan *Variance Inflation Factors* (*VIF*). Jika nilai dari *VIF* mendekati 1 berarti tidak terjadi multikolinearitas, namun bila lebih besar dari 10 maka menunjukkan bahwa terjadi multikolinearitas yang serius. Untuk data Indeks Pembangunan Manusia berikut nilai *VIF* untuk setiap peubah bebas.

Tabel 1. Analisis Nilai *VIF* dengan Peubah $X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}, X_{i6}, X_{i7}$ dan X_{i8}

Peubah Bebas	***VIF***
X_{i1}	4,7
X_{i2}	**14,5**
X_{i3}	**34,5**
X_{i4}	1,3
X_{i5}	**14,8**
X_{i6}	2,1
X_{i7}	1,1
X_{i8}	1,7

Dari hasil *Output* Minitab 14.0 untuk *Variance Inflation Factors* (*VIF*) yang ditunjukkan pada Tabel 1, terlihat bahwa pada X_{i2}, X_{i3} dan X_{i5} terjadi multikolinearitas.

Penaksiran parameter pada data Indeks Pembangunan Manusia menggunakan metode Kuadrat Terkecil akan memperoleh hasil sebagai berikut :

$\hat{\beta}_0 = 20{,}7$ $\qquad$ $\hat{\beta}_4 = 0{,}408$ $\qquad$ $\hat{\beta}_8 = 0{,}0713$

$\hat{\beta}_1 = 0{,}153$ $\hat{\beta}_5 = -0{,}109$ $\hat{\beta}_2 = -0{,}030$

$\hat{\beta}_6 = 0{,}0046$ $\hat{\beta}_3 = 0{,}321$ $\hat{\beta}_7 = -0{,}076$

Dengan demikian data Indeks Pemabangunan Manusia mempunyai model lengkap

$$\hat{Y}_i = 20{,}7 + 0{,}153X_{i1} - 0{,}030X_{i2} + 0{,}321X_{i3} + 0{,}408X_{i4} - 0{,}109X_{i5} + 0{,}0046X_{i6} - 0{,}076X_{i7} + 0{,}0713X_{i8}$$

4.2. Perbandingan Metode Himpunan Bagian Terbaik

Metode Himpunan Bagian Terbaik

Semua kemungkinan yang telah ditemukan dengan menggunakan metode *All Possible Regression*, kita dapat menggunakan metode Himpunan Bagian Terbaik untuk dapat menghitung hanya sebagian dari semua kemungkinan regresi. Dalam hal ini pengolahan data dengan metode Himpunan Bagian Terbaik menggunakan program Minitab 14.0.

Tabel 2. Nilai R-Sq, R-Sq (adj), Cp Mallow, dan S.

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	X 1	X 2	X 3	X 4	X 5	X 6	X 7	X 8
1	78,8	77,9	96,7	1,7060			X					
1	72,6	71,4	131,2	1,9400					X			
2	91,0	90,2	31,0	1,1375			X	X				
2	84,1	82,6	69,5	1,5130				X	X			
3	95,8	95,2	6,3	0,79418	X		X	X				
3	94,3	93,5	14,5	0,92387			X	X	X			
4	96,3	95,6	5,4	0,76164			X	X	X			X
4	96,2	95,4	6,4	0,77945	X	X	X	X				
5	97,0	96,3	3,5	0,70188	X		X	X	X			X
5	96,7	95,8	5,4	0,74190	X	X	X	X				X
6	97,1	96,1	5,1	0,71223	X		X	X	X		X	X
6	97,1	96,1	5,4	0,71943	X	X	X	X	X			X
7	97,1	95,9	7,0	0,73173	X	X	X	X	X		X	X
7	97,1	95,9	7,0	0,73276	X		X	X	X	X	X	X
8	**97,1**	**95,7**	**9,0**	**0,75414**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Pemilihan model terbaik dengan menggunakan metode Himpunan Bagian Terbaik berdasarkan kriteria $Cp\ Mallows$ yang mendekati p parameter seperti yang terlihat pada Tabel 2 menghasilkan model regresi:

$$y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \varepsilon_i$$

Dengan $R^2_{adj} = 95{,}7\%$, dimana 4,3% dipengaruhi oleh variable lain. Pada model regresi diatas nilai *VIF* untuk menunjukkan nilai multikolinearitas di atas. Untuk *Variance Inflation Factors* (*VIF*) terlihat bahwa pada X_{i2}, X_{i3} dan X_{i5} terjadi multikolinearitas. Selanjutnya adalah menaksir parameter dengan menggunakan metode Kuadrat Terkecil. Dengan demikian, data Indeks Pembangunan Manusia bila dimodelkan menggunakan metode Himpunan Bagian Terbaik akan menghasilkan model sebagai berikut :

$$\hat{Y}_i = 20{,}7 + 0{,}153X_{i1} - 0{,}030X_{i2} + 0{,}321X_{i3} + 0{,}408X_{i4} - 0{,}109X_{i5} + 0{,}0046X_{i6} - 0{,}076X_{i7} + 0{,}0713X_{i8}$$

Metode Fraksi

Seperti pada metode Himpunan Bagian Terbaik, data yang digunakan pada contoh penerapan metode fraksi adalah data Indeks Pembangunan Manusia. Peubah terikat (Y_i) adalah Indeks Pembangunan Manusia dan peubah bebasnya (X_{i1}) Indeks komponen kesehatan; (X_{i2}) Indeks angka melek huruf; (X_{i3}) Indeks pendidikan; (X_{i4}) Indeks daya beli; (X_{i5}) Pengeluaran rata-rata per kapita untuk sub golongan makanan; (X_{i6}) Tingkat pengangguran terbuka; (X_{i7}) Jumlah pemuda (penduduk usia 15-35 tahun) menurut Kabupaten/Kota.

Pada metode Fraksi peubah-peubah bebas di kelompokkan menurut tingkat korelasinya satu sama lain. Tingkat korelasi antar peubah bebas maupun tingkat korelasi antara peubah terikat dengan peubah bebas. Korelasi peubah terikat dan peubah bebas dapat diketahui dengan menggunakan bantuan Microsoft Excel dan program Minitab 14.0.

Tabel 3. Korelasi Data Indeks Pembangunan Manusia.

Peubah Bebas	Korelasi							
	Y_i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	X_{i6}	X_{i7}
X_{i1}	0,787							
X_{i2}	0,674	0,421						
X_{i3}	0,888	0,704	0,880					
X_{i4}	0,572	0,204	0,168	0,266				
X_{i5}	-0,852	-0,830	-0,483	-0,808	-0,292			
X_{i6}	0,419	0,396	0,301	0,4883	-0,083	-0,583		
X_{i7}	-0,077	0,080	-0,153	-0,106	-0,025	-0,037	-0,009	
X_{i8}	0,255	0,282	0,309	0,279	-0,129	-0,059	0,235	-0,147

Berdasarkan Tabel 3 dapat dilihat bahwa Y_i berkorelasi dengan X_{i1}, X_{i2}, X_{i3} dan X_{i5} yaitu berturut-turut 0,787 , 0,674 , 0,888 dan -0,852. Peubah X_{i1}, X_{i3} dan X_{i5} saling berkorelasi kuat. Sementara X_{i2} berkorelasi kuat dengan X_{i3}. X_{i4} berkorelasi dengan Y_i namun tidak cukup kuat. Peubah bebas X_{i6}, X_{i7} dan X_{i8} tidak berkorelasi dengan peubah bebas lainnya, dimana dapat dikelompokkan sebagai berikut :

1. X_{i1}, X_{i3} dan X_{i5}
2. X_{i2} dan X_{i3}
3. X_{i4}
4. X_{i6}, X_{i7} dan X_{i8}

Setelah peubah bebas dikelompokkan sesuai dengan tingkat korelasinya, selanjutnya akan dipilih fraksi ½ untuk dapat menentukan jumlah peubah bebas yang akan dimasukkan kedalam model. Dengan memilih fraksi ½ berarti akan ada 4 peubah bebas yang akan masuk dalam model.

Dari pengelompokkan di atas akan dipilih peubah bebas dari masing-masing kelompok sesuai dengan tingkat korelasinya pada Y_i. Dari X_{i1}, X_{i3} dan X_{i5} dipilih X_{i3} dan X_{i5} karena berkorelasi paling kuat dengan Y_i. Jadi akan dipilih 2 dari peubah X_{i4}, X_{i6}, X_{i7} dan X_{i8} dipilih X_{i4} dan X_{i6} karena memiliki korelasi yang paling kuat terhadap Y_i. Jadi model yang dipilih mengandung peubah-peubah X_{i3}, X_{i4}, X_{i5} dan X_{i6}

$$Y = \beta_0 + \beta_3 X_{i3} + \beta_4 X_{i4} - \beta_5 X_{i5} + \beta_6 X_{i6} + \varepsilon_i$$

Dengan $R^2_{adj} = 93{,}2\%$ dimana 6,8% dipengaruhi oleh variabel lain. Pada model regresi diatas nilai *VIF* menunjukkan tidak terjadi multikolinearitas. Setelah melihat nilai *VIF* maka selanjutnya adalah menaksir parameter sehingga diperoleh model regresi jika menggunakan metode fraksi yaitu :

$$\widehat{Y_i} = 28{,}1 + 0{,}354 X_{i3} + 0{,}378 X_{i4} - 0{,}161 X_{i5} + 0{,}0021 X_{i6}$$

Meskipun pada model regresi dengan menggunakan fraksi ½ masih ada peubah bebas yang tidak nyata tetapi model ini memiliki $R^2_{adj} = 93{,}2\%$ dan tidak ada masalah kolinearitas lagi. Jika model ini dianggap kurang memuaskan karena ada peubah yang tidak nyata maka dapat digunakan fraksi lebih kecil yaitu ¼ atau dengan 2 peubah bebas. Metode fraksi akan memilih peubah X_{i3} yang berkorelasi paling kuat dengan Y_i dan peubah bebas X_{i4} yang berkorelasi nyata dengan Y_i namun tidak berkorelasi dengan X_{i3}. Analisis terhadap model dengan peubah bebas X_{i3} dan X_{i4} adalah

$$\widehat{Y_i} = 4{,}73 + 0{,}514 X_{i3} + 0{,}407 X_{i4}$$

Dengan nilai *VIF* sebesar 1,1 untuk kedua peubah bebas dan $R^2_{adj} = 90{,}2\%$ dimana 9,8% dipengaruhi oleh model lain, model ini sangat baik untuk menjelaskan Y_i.

5. Kesimpulan

Berdasarkan uraian pembahasan mengenai metode Himpunan Bagian Terbaik dan metode fraksi dalam pemilihan model regresi berganda terbaik, maka dapat diambil kesimpulkan sebagai berikut:

1. Data Indeks Pembangunan Manusia bila dimodelkan menggunakan metode Himpunan Bagian Terbaik akan menghasilkan model sebagai berikut :

$$\hat{Y} = \beta_0 + \beta_1 X_{i1} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta \quad X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \varepsilon_i$$

2. Pada metode Fraksi peubah-peubah bebas di kelompokkan menurut tingkat korelasinya satu sama lain. Tingkat korelasi antar peubah bebas maupun tingkat korelasi antara peubah terikat dengan peubah bebas. Model regresi jika menggunakan metode fraksi yaitu :

$$\widehat{Y_i} = 28{,}1 + 0{,}354 X_{i3} + 0{,}378 X_{i4} - 0{,}161 X_{i5} + 0{,}0021 X_{i6}$$

3. Pemilihan model regresi berganda terbaik dengan menggunakan metode Himpunan Bagian Terbaik kurang valid untuk data multikolinearitas. Sehingga digunakan metode

Fraksi yang lebih fokus pada masalah adanya multikolinearitas, sehingga menghasilkan model yang lebih valid.

Seperti yang telah dijelaskan metode fraksi mengelompokkan peubah-peubah pada model lengkap berdasarkan tingkat korelasinya kemudian mulai memilih fraksi yang akan digunakan untuk dapat mengetahui jumlah peubah bebas yang akan dimasukkan kedalam model guna memperoleh model regresi berganda terbaik. Pada metode fraksi terdapat keharusan untuk memuat fraksi dari jumlah peubah bebas dalam data yang menjadi penghambat untuk lebih bebas menentukan model terbaik. Penulis menyarankan untuk selalu membandingkan metode fraksi dengan metode-metode lain seperti Himpunan Bagian Terbaik atau metode *Stepwise*.

Daftar Pustaka

[1] Draper N., and Smith H., 1992. *Analisis Regresi Terapan Edisi Kedua* (Terjemahan oleh Bambang Sumantri). Gramedia, Jakarta.

[2] Hanum H., 2009. Penentuan Model Regresi Berganda Terbaik Menggunakan Metode Fraksi. *Penelitian Fundamental,* Universitas Sriwijaya.

[3] Karim M.E., 2012. Selecting of The Best Regression Equation by Sorting Out Variables. *hhtp://www.angelfire.com/ab5/get5/selreg.pdf* [diunduh 3 Februari 2012].

[4] O’Brien R.M., 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41:673-690.

[5] Tiro M.A, 2000. *Analisis Korelasi dan Regresi Edisi Kedua*. Makassar State University Press, Makassar.